Int J Life Cycle Assess (2014) 19:1733–1744
DOI 10.1007/s11367-014-0771-9

LCI METHODOLOGY AND DATABASES

Revisiting least squares techniques for the purposes of allocation in life cycle inventory

Nathan B. Cruze · Prem K. Goel · Bhavik R. Bakshi

Received: 24 June 2013 / Accepted: 20 June 2014 / Published online: 5 August 2014

Abstract
Background, aim, and scope In this communication we reexamine the use of various least squares techniques (namely ordinary least squares (OLS), data least squares (DLS), and total least squares (TLS)) for the purposes of allocation as proposed in an earlier article in this journal (Marvuglia et al. (Int. J. Life Cycle Assess. 15:1020-1040, 2010)). These methods are placed within the context of traditional methods of partitioning allocation. An equivalence between least squares techniques and traditional partitioning is noted and demonstrated on previously published brick production data.
Materials and methods A short summary of the relevant least squares techniques is provided followed by a description of the problem of inventory calculation for the case of more products than processes. This is presented in terms of non-unique solutions to underdetermined systems of linear equations with intensity matrix as unknown.
Results and discussion We provide another analysis of the Sicilian brick production case study. Upon reexamination of the brick data, a number of disparities in the published inventories for brick can be more fully explained in terms of (1) data quality, (2) additional assumptions made to extract a solution of an underdetermined system of equations, and (3) discrepancy vectors.
Conclusions and outlook Like other types of partitioning, inventories produced by least squares techniques correspond to one of the non-unique solutions of an underdetermined system of linear equations. Based on this insight, we advise against the use of least squares techniques as a black-box approach to the allocation problem and conclude that the recommendation of TLS on the basis of its asymptotic properties is not theoretically justified.

Keywords Allocation · Least squares · Matrix-based LCI · Partitioning · Regression

1 Introduction

A source of subjectivity in the inventory phase, allocation has been called one of the Achilles heels in the practice of life cycle assessment (Marvuglia et al. 2010, p. 1036). The life cycle inventory literature contains a variety of proposals for carrying out allocation, and Cherubini and Strømman (2011) noted that expert preferences appear to be evenly distributed across the range of methods. Wardenaar et al. (2012) observed that a number of existing environmental protocols actually recommend different methods of allocation during the inventory phase and debated the implications for policy applications.

Recently, Marvuglia et al. (2010, 2012) have proposed the use of several least squares techniques for the purposes of allocation and demonstrated their calculation on data representing a brick production network in Sicily. One part of the appeal of the least squares techniques seemed to be that the scale factors were unique or "optimal" approximate solutions of an overdetermined system of linear equations with respect to some loss or objective function, perhaps eliminating some subjective elements in the inventory calculation. They ultimately recommended a total least squares (TLS) approach on the basis of asymptotic properties (i.e., strong consistency)

Responsible editor: Nydia Suppen-Reynag

Electronic supplementary material The online version of this article (doi:10.1007/s11367-014-0771-9) contains supplementary material, which is available to authorized users.

N. B. Cruze · P. K. Goel
Department of Statistics, Columbus, USA

B. R. Bakshi (✉)
Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210, USA
e-mail: bakshi.2@osu.edu

of the TLS estimator for scale factors. Regarding the relationship between least squares methods and typical partitioning, they opined:

> "One could argue that solving the inventory problem directly in its rectangular form (without making the system matrix square) intrinsically leaves unallocated the single functions of a multi-functional process without actually solving the problem of allocation, but just hiding it. However, in our view, making resort to allocation has no other usefulness than allowing the transformation of a rectangular system into a square one". (Marvuglia et al. 2010, p. 1024)

In the next section, we summarize earlier work on the subject of partitioning of a matrix-based life cycle inventory (Cruze et al. (2013b)) which presented partitioning allocation as a particular solution of an underdetermined system of linear equations. Allocation is an attempt to state life cycle intensities of one of several products given fewer observations than unknowns. From this viewpoint, inventories derived by least squares techniques are not unique solutions. Rather, they are a just a few among infinitely many solutions to an underdetermined system of linear equations. Furthermore, least squares techniques are not something done in lieu of partitioning. They are equivalent to partitioning, sometimes with weights that don't make practical sense. These claims are illustrated through reexamination of the original brick data in Marvuglia et al. (2010). We conclude in Section 4 that least squares techniques are not a black box solution to the allocation problem.

2 Inventory calculation and allocation

In our notation throughout this paper, a bold capital letter will denote a matrix. Bold lowercase letters denote column vectors, and an italic variable denotes a scalar value. Let n denote the number of processes, p denote the number of products, and q denote the number of distinct interventions included in the study. Partitioning in matrix LCI and the least squares techniques deals specifically with the case $n<p$, therefore at least one multi-output process must be present.

A fundamental relation of life cycle inventory describes the net outputs of a system $\boldsymbol{f}_{p\times 1}$ in terms of the scale $\boldsymbol{s}_{n\times 1}$ at which each process (column of the technology matrix $\boldsymbol{A}_{p\times n}$) must operate:

$$As = f. \tag{1}$$

One may wish to know at what scale each process must be operated to provide some quantity of the j^{th} product, the functional unit $\boldsymbol{f}_j$. Provided that the technology matrix is square with $p=n$ and $\boldsymbol{A}$ invertible, the corresponding list of scale factors $\boldsymbol{s}_j$ may be found straightforwardly by inversion:

$$s_j = A^{-1} f_j. \tag{2}$$

If total interventions at each process are in direct proportion to the scale of operation, the life cycle inventory of the j^{th} product, $\boldsymbol{g}_j$ is given by the inner product of the interventions matrix $\mathbf{B}_{q\times n}$ with the vector of scale factors as follows:

$$g_j = Bs_j \tag{3}$$

For the invertible technology matrix, Heijungs and Suh (2002) define the intensity matrix as

$$\Lambda = BA^{-1}. \tag{4}$$

The inventory for one unit of the j^{th} product, $\boldsymbol{g}_j$, is the j^{th} column of $\boldsymbol{\Lambda}_{q\times p}$. Trivially, the intensity matrix Λ must satisfy

$$A^{\mathrm{T}} \Lambda^{\mathrm{T}} = B^{\mathrm{T}}. \tag{5}$$

When there are more products or functions present than processes, Eq. 1 is an overdetermined system of equations, and it may not have an exact solution. This marks the point of departure for much of the allocation literature. Many methods of partitioning have been developed in which multifunctional processes are broken into virtual, single-function processes to obtain an invertible technology matrix $\boldsymbol{A}_*$ (of dimension $p\times p$), and shares of inputs and interventions are assigned in a partitioned interventions matrix $\boldsymbol{B}_*$ ($q\times p$) in proportion to some characteristic of the joint products (e.g., mass, economic value, etc.) so that the final inventory $\boldsymbol{g}_*$ may be calculated as

$$g_* = B_* A_*^{-1} f. \tag{6}$$

2.1 Three least squares methods

Marvuglia et al. sought a solution of the overdetermined Eq. 1 without having to explicitly define some new technology matrix $\boldsymbol{A}_*$ and interventions matrix $\boldsymbol{B}_*$. They proposed ordinary least squares (OLS) allocation, data least squares (DLS) allocation and total least squares (TLS) allocation as three approaches to obtain an approximate solution (vector of scale factors) to Eq. 1. All three types of least squares problem can be written compactly as variants of the TLS minimization problem (DeGroat and Dowling (1993); Van Huffel and Vandewalle (1991); Van Huffel (1997)):

$$\begin{aligned} & min_{s,\Delta A,\Delta f} \| [\Delta A | \Delta f] \|_{F} \\ & subject\ \ to\ \ (A + \Delta A)s = f + \Delta f. \end{aligned} \tag{7}$$

The OLS result may be obtained by setting $\boldsymbol{\Delta A=0}$, while the DLS result is obtained by setting $\boldsymbol{\Delta f=0}$ in Eq. 7 and then solving the programming problem (DeGroat and Dowling (1993)). In words, these programs seek a matrix of the smallest corrections to the technology matrix and functional unit, denoted $[\boldsymbol{\Delta A}|\boldsymbol{\Delta f}]$, in the sense that its Frobenius norm, denoted $||\bullet||_{F,}$ is minimal. Under these minimal changes, the input-output relationships of the perturbed system shown in the constraints of Eq. 7 may be satisfied exactly. Given scale factors as unique minimizers of Eq. 7, the inventory vector follows the computation shown in Eq. 3 since the scale factors are obtained in the dimension of processes. Since the three minimization problems differ from one another in terms of their objective functions and constraints, the scale factors and resulting inventories produced by the three least squares methods need not be identical.

2.2 Allocation and the underdetermined system of linear equations

When the number of products in the system is greater than the number of processes, at least one process is multifunctional, and Eq. 5 represents an underdetermined system of equations, with the number of equations equal to the number of processes contained in the technology matrix. Without loss of generality, one can consider just one type of intervention at a time so that

$$A^{\mathrm{T}} r_i = \ b_i, \tag{8}$$

where the vectors $\boldsymbol{r_i^{\mathrm{T}}}$ and b_i^T are the i^{th} *rows* of the intensity matrix Λ and interventions matrix $\boldsymbol{B}$, respectively.

When finding a solution via partitioning, the system of equations

$$A_*^T r = b_* \tag{9}$$

is solved by inversion to obtain a unique intensity vector, say $\boldsymbol{r}_*$, and the linear combination $\boldsymbol{f^T r_*}$ is interpreted as a component of the inventory related to the functional unit. While the solution to Eq. 9 is unique, Eq. 8 has infinitely many exact solutions. The solutions to Eq. 9 and Eq. 8 are closely related, the unique solution of the former being just one particular solution of the latter.

More in-depth study of the equivalence of inventories derived from those systems of equations is given in Cruze et al. (2013b), but we wish to summarize the following here:

1. A partitioning allocation as in Eq. 9 obtains one solution of Eq. 8 made under additional equality constraints called *side conditions*,
2. Obtaining the solution under side conditions is equivalent to solving Eq. 8 by some generalized inverse of the technology matrix, $\boldsymbol{A^{T.}}$ The meaning of these two points may be made more concrete by study of OLS allocation since the generalized inverse in point 2 is explicitly known. The Moore-Penrose pseudoinverse, $(\boldsymbol{A^T})^+$ is used to solve Eq. 8 so that

$$r_{\mathrm{OLS}} \ = \ \left(A^T\right)^+ b. \tag{10}$$

Taken together with point 1, there is an equivalent partitioning of the technology and intervention matrices so that r_{OLS} can be obtained from the operations in Eq. 9.

Similarly, equivalent partitionings can be found which contain DLS and TLS intensities for the target product in the resulting intensity vectors. The upshot is that the least squares techniques pursued by Marvuglia et al. are not something done in lieu of partitioning; they are each equivalent to partitioning. Moreover, they are not unique solutions in the sense of Eq. 8 despite the fact that the vectors of scale factors may be optimal in the sense of the programming problem stated in Eq. 7. There is no reason to expect that a least squares solution of Eq. 7 produces an inventory that is more physically meaningful, interpretable or actionable than more standard types of partitioning. Indeed, the debate as to which method of allocation is “right” persists because in practice, different allocations entail specific constraints on some parameters of the intensity vector in Eq. 8, and yet the outcome is interpreted without regard to those constraints.

3 The brick production study revisited

The raw, unallocated data for the Sicilian brick production system may be found in Marvuglia et al. (2010). The original process technology matrix consisted of 8 processes which exchanged 14 economic flows. Six multifunctional processes were present, with recycled inerts being an additional joint output of the brick production process. In addition, 21 types of environmental interventions were considered; these included several types of natural resources consumed as well as emissions to air and water, and three wastes.

Marvuglia et al. compared five types of allocations. Three inventories were based on the selection of an approximate solution of Eq. 1 by OLS, DLS, and TLS regression-based techniques applied to the 14×8 system. The other two partitioning-based allocations were derived from technology matrices with dimensions of 13 products by 13 virtual processes. One of these allocations was based on physical considerations while the other was a partitioning allocation based

on the economic valuation of the joint products. The reduction in number of products in these traditional methods of allocation came from a partial substitution of recycled inerts for sand.

The disparities across allocations were significant. A few of the many disparities shown in the original tables (Marvuglia et al. 2010, Table 11) are summarized as follows:

- The life cycle releases of emissions and wastes, and uses of resources assigned to 1 t of bricks by the OLS allocation are all minuscule in comparison to any of the other allocation procedures, the largest releases and consumption being measured on the order of joules or grams as opposed to tens or hundreds of megajoules or kilograms. We note that this is due to the choice of the Moore-Penrose inverse to solve Eq. 8, *which produces the shortest length solution among all possible solutions of an underdetemined system of linear equations* (Golub and Kahan 1965, p. 205).
- By convention, coal consumption (a resource input from outside the system boundary) should be negative, yet all five methods produce positive results spanning ten orders of magnitude, with OLS allocation on the order of 5 J, physical allocation at *856 megajoules* and economic allocation on the order of *12,200 megajoules*!
- Water requirements are of the same order of magnitude under physical and economic allocations, but of *opposite signs*, as are life cycle emissions of CO_2, CO, and non-methane volatile organic compounds (*NMVOC*).
- Marvuglia, Cellura and Heijungs observed similarities between physical allocation and DLS and TLS allocations for a select few emissions, wastes, and resources and concluded that this lent some credence to the TLS technique, but even the two traditional allocations agreed closely on emissions of phosphorus, nitrogen, and adsorbable organohalides (*AOX*).

It is important to understand these disparities with regard to possible issues of data quality, the additional conditions imposed by partitioning and least squares allocation, and the provided discrepancy vectors.

3.1 Inventory disparities and data quality

The brick production process required inputs of virtually all other products in the system. The network featured seven upstream processes: production of crude oil, oil-derivatives, natural gas, electricity and heat, clay, sand and gravel, and biomass. These processes were the unique producers of all the necessary inputs within the system boundary; however, they do not satisfy system-wide needs operating together at positive scale in the unallocated matrix. This relates specifically to energy inputs originating in the heat and electricity, natural gas, and oil derivative processes. If the assumption of linearity is to be believed, then these processes cannot be operated together at positive scale to satisfy their own needs, much less to satisfy needs of other processes in the network. The processes upstream of brick are the unique producers of their respective materials or energy inputs within the system boundary, but the *unallocated* electricity and heat, oil derivatives and natural gas processes cannot be rescaled to operate together in the forward direction as follows:

1. The electricity/heat process requires 4.27 MJ natural gas to produce its joint outputs; the natural gas process in Table 1 has been rescaled accordingly.
2. The natural gas process requires 154 MJ of diesel oil to operate at that level, the oil derivatives process shown in Table 1 has rescaled accordingly.
3. The combined needs of the oil derivatives and natural gas processes exceeds the heat and electricity resources available, as indicated by the negative values in the list of the net outputs for these three processes.

The negative energy values indicate a shortfall of electricity and heat inputs going to the oil derivatives and natural gas processes. An even larger shortfall occurs when one attempts to couple the five remaining processes with these three. This suggests that:

- the provided unit processes may be poorly representative of actual processes involved in the network, or
- that some important energy processes may have been overlooked within the system boundary, or
- if the measurements on each process are accurate, the phenomenon is poorly described by a linear model.

This apparent shortfall of energy inputs could arise simply if data for one or more of the obtained processes were poor representatives of actual processes used in the Sicilian network or if some other gross error was present. However, it has far reaching implications for all of the calculated inventories.

We note that several of the originally tabulated inventories (Marvuglia et al. 2010, Table 11) failed to maintain sign convention. As discussed above, all five inventories showed that the natural resource coal was a positive quantity, indicating that it had been produced or otherwise its use was somehow avoided on net by the system. The possibility of sign change in the final inventory can be attributed to the existence of one or more negative scale factors. The originally published scale factors for each method of allocation attempted in (Marvuglia et al. 2010, Table 10) show a mix of positive and

Table 1 A physical inconsistency in a submatrix of the original technology matrix. The result is a system wide shortfall of energy inputs throughout the network

		Processes			Net output
		Electricity/heat	Oil derivatives	Natural gas	
Products	Electricity (MJ)	1	−4.93E+00	0	−3.93E+00
	Heat (MJ)	2.48E+00	−1.74E+02	−2.08E-01	−1.72E+02
	Diesel oil (MJ)	0	1.54E+02	−1.54E+02	0
	Fuel oil (MJ)	0	1.54E+02	0	1.54E+02
	Natural gas (MJ)	−4.27E+00	0	4.27E+00	0

negative signs. *We note that in all five cases, the negative signs correspond to processes (or virtual processes for physical and economic allocation) which relate to heat, electricity, natural gas, and the oil derivatives.*

In system boundary expansion, negative scale factors are sometimes interpreted as an avoided activity or a volume of activity that has been displaced by other production. In this particular case, the processes associated with negative scale factors are not additional processes that might be offset by production, but rather *the unique producers of their respective energy types within the system boundary.* The existence of negative scale factors in this example runs counter to a claim made in Heijungs and Tan (2010) that an inverse-positive technology matrix (one in which $\boldsymbol{A}_{*}^{-1}\geq\boldsymbol{0}$ holds elementwise) could be obtained by substitution and partitioning assumptions, and that it would preclude the use of structural path analysis techniques a la Strømman et al. (2009), results on fuzzy methods for uncertainty propagation (Tan (2008); Heijungs and Tan (2010); Cruze et al. (2013a)), and the Monte-Carlo algorithm of Peters (2007).

These distinctions are illustrated in Figs. 1 and 2; Tables 4 and 5 in Section 1.2 of the supporting information were used to derive these figures. Each of these figures represents activity within the rescaled virtual network for physical allocation (Fig. 1) and a slight variation of the physical allocation which sets $w_{\text{diesel oil}}=0.01$ while keeping all other four allocation weights fixed (Fig. 2). Selected interventions at each virtual process are also shown; the reported life cycle intensities are the sums of these interventions at all 13 virtual process. Line thickness indicates the order of magnitude of each flow. The virtual processes shown in red in Fig. 1 are associated with a negative scale factor. A literal interpretation of the physical allocation would suggest that not only is coal a net product of the system, some virtual processes make multiple products, and these products originate in unexpected places, e.g., the diesel virtual process produces heat, electricity and crude oil, and seems to draw diesel oil from the natural gas process. Directed graphs for each of the five different inventories presented in Marvuglia et al. (2010) would be similar, and each precludes the use of some of the aforementioned analytical techniques for error propagation and inventory hybridization. By contrast, an inverse-positive outcome like that shown in Fig. 2 maintains sign of all inventories, and each virtual process serves just a single function as expected.

Using MATLAB's powerful symbolic features (see Section 1.2 of the supporting information for more details), we determined that inverse-positivity, a property which would at least preserve the signs of all final inventories, depends exclusively on the pair of weights assigned to electricity and diesel oil, denoted ($w_{\text{electricity}}$, $w_{\text{diesel oil}}$), in a conventional partitioning of the brick data. Keeping weights assigned to three other virtual processes ($w_{\text{white clay}}$, w_{sand}, $w_{\text{olive cake}}$) fixed at their physical allocation values, the projection shown in Fig. 3 illustrates that as the threshold is approached from either side, the life cycle intensity of coal can be driven to ±∞ since the allocated technology matrix approaches singularity for these allocations. Only ($w_{\text{electricity}}$, $w_{\text{diesel oil}}$) pairs on the side of the threshold corresponding to small values of $w_{diesel\ \text{oil}}$ produce inventories of coal which conform to sign assumptions.

As we discuss in the next section, inventories calculated by several methods of allocation, even inverse-positive outcomes, may differ anyway as each solution of Eq. 8 is derived under a different set of additional constraints. In this example, however, *data quality* may have added to the disparities among inventories since the inventories are not even bounded and may change sign. This explains why the economic and physical allocations often differed by orders of magnitude in the originally tabulated inventories. One key lesson this example provides is that none of the presented partitionings, least squares allocations included, can overcome a data quality issue. If it is known that all elements of the published technology matrix are accurate so that it is indeed a legitimate technology matrix, then this should lead us to question assumptions of linearity, in which case existing partitioning and least squares methods may not even be pertinent.

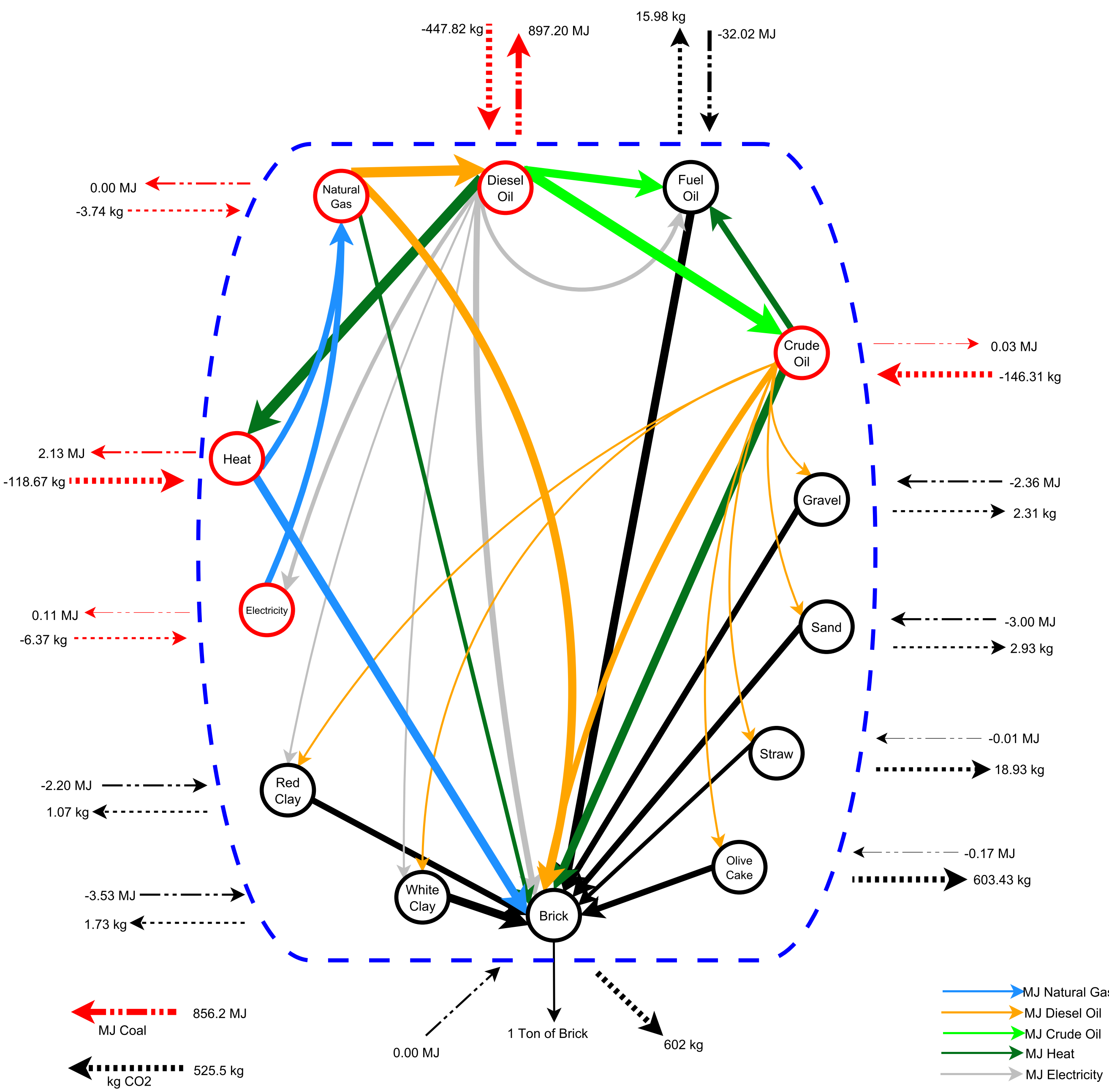

Fig. 1 Rescaled virtual network for physical allocation. When the five energy processes are rescaled to reflect negative scale factors, the structure of the network is changed so that after partitioning some processes may appear to serve more than one function and some processes make the same type of product

3.2 Additional assumptions and least squares equivalent partitionings of the brick data

Although derived from believable information, it is understood that as a representation of any actual production structure, the technology matrix after allocation is an imaginary creation. Joint products of the same process may not be produced independently, and inputs, resources, emissions and wastes assigned to the virtual, single-function processes vary seemingly arbitrarily from one choice of partitioning to the next. Marvuglia et al. had hoped that some of the more subjective elements of allocation might be eliminated through the use of various least squares optimality criteria.

Two comments are worth making here:

- First, since allocation is simply a matter of finding one particular solution to Eq. 8, inventories reached by traditional procedures or by least squares techniques are both made under a set of equality constraints on non-identifiable intensities. Since the data contain no

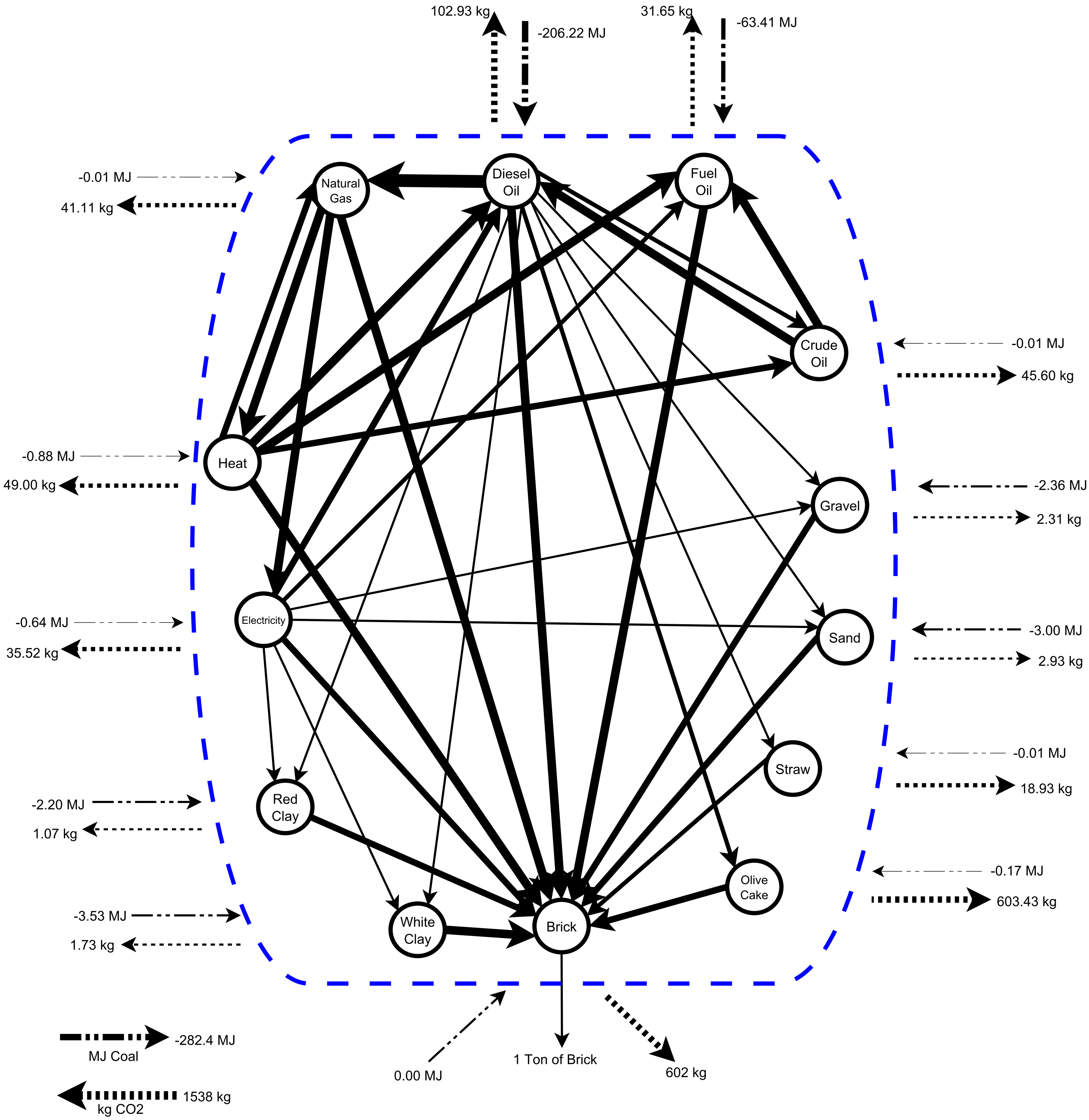

Fig. 2 Rescaled virtual network for an inverse-positive allocation. In this case, each virtual process serves a single function and is the unique producer of its respective energy source or material

information about these constraints, they can be established arbitrarily with far-reaching implications for the interpretation of the inventory vector.

- As motivated in Section 2.2 the least squares procedures do not necessitate the actual construction of square matrices, but they are equivalent to doing so, sometimes with weights that correspond to indefensible allocations, allocations with corresponding weights that sum to 1, but fall outside the interval [0, 1].

Both of these points are further explicated below.

3.2.1 Side conditions

For the brick data, the corresponding Eq. 8 is a system of 8 equations in the 14 unknown life cycle intensities for each product, none of which are uniquely identified in the data at hand. The derivation of these additional conditions, or side conditions, from the allocated technology matrix is discussed

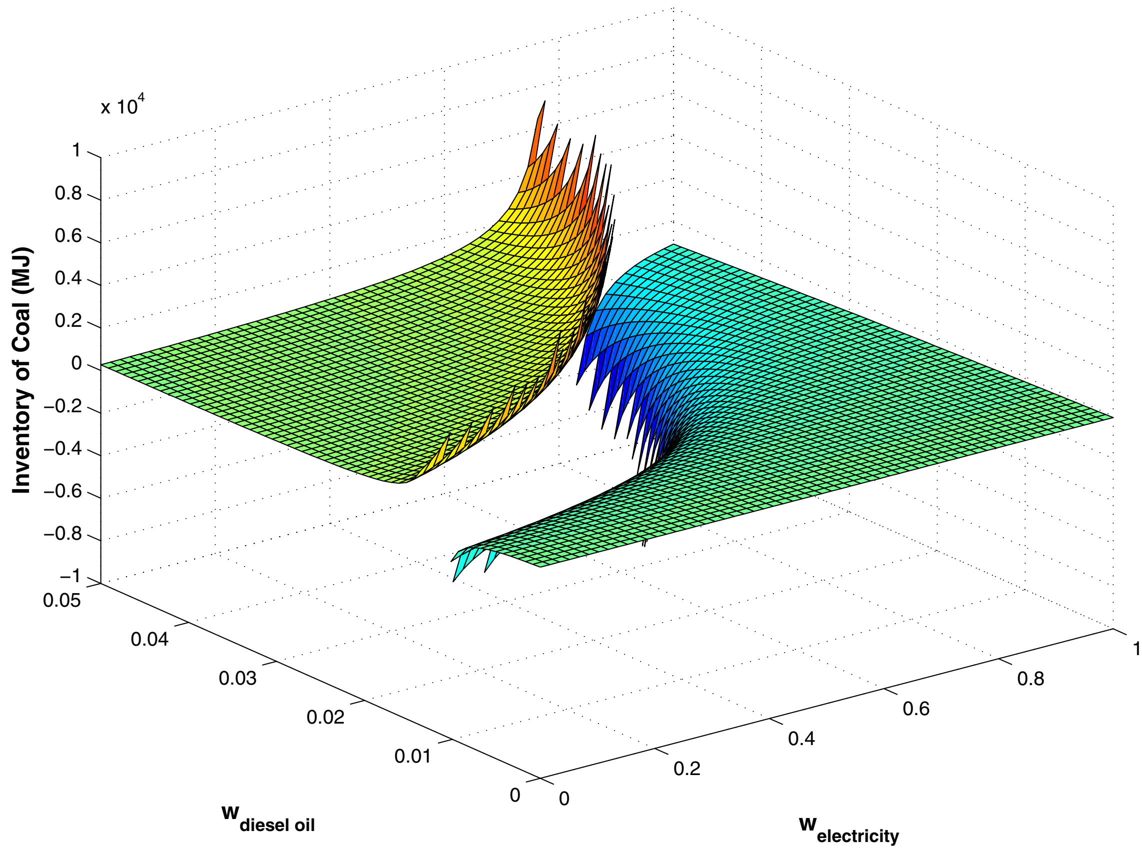

Fig. 3 Life cycle intensity of coal associated with 1 t of brick as a function of allocation weights $w_{\text{electricity}}$ and $w_{\text{diesel oil}}$. In this figure, the remaining allocation weights have been fixed at $w_{\text{white clay}}=w_{\text{sand}}=0.5$ and $w_{\text{olive cake}}=0.8$ in keeping with allocation weights for physical allocation

in more detail in Cruze et al. (2013b), but for illustrative purposes, we present one set of additional constraints imposed by physical allocation to obtain an intensity vector with respect to coal consumption:

$$r_{\text{inerts}} = .85r_{\text{sand}} \tag{11}$$

$$r_{\text{electricity}}-2.48r_{\text{heat}} = 2.562\mathrm{r}_{\text{natural gas}}-0.0009 \tag{12}$$

$$r_{\text{white clay}} = r_{\text{red clay}} \tag{13}$$

$$r_{\text{sand}} = r_{\text{gravel}} \tag{14}$$

$$r_{\text{olive cake}}-r_{\text{straw}} = 0.2436r_{\text{diesel oil}}-0.0008 \tag{15}$$

$$r_{\text{fuel oil}} = r_{\text{diesel oil}}. \tag{16}$$

These side conditions represent several assumed relationships about life cycle coal consumption among the various products included within the network boundaries. To reduce the dimension of products, Marvuglia et al. made an assumption that only 85 % of recycle inerts could be substituted for sand. This is also a statement about the *relative impacts* of the two materials (Eq. 11). Physical allocation suggests that coal intensities between some pairs of joint products are indistinguishable, namely that the effects per kilogram of either clay are the same (Eq. 13), inputs of coal into sand and gravel production are identical (Eq. 14), and the life cycle consumption of coal by both oil derivatives is the same (Eq. 16). In Eqs. 12 and 15, the relative magnitudes of life cycle coal use also depend on the inputs into those two processes. These equations are merely one possible collection of six additional assumptions under which all 14 product intensities may be obtained. Table 6 provided in the supporting information describes the entire family of allocation equivalent constraints in terms of 85 % recovery rate of inerts, the five allocation weights for other multifunctional processes, and any type of 21 given environmental interventions.

For a comparison, Table 2 shows intensity vectors $\boldsymbol{r}_{\text{phys.}}$, $\boldsymbol{r}_{\text{econ.}}$, and $\boldsymbol{r}_{\text{OLS}}$. (These are the first rows of their

Table 2 Individual product inventories of life cycle coal (MJ) for brick production network dictated under the side conditions of physical, economic, and OLS allocations. Negative sign indicates net use by a product, positive sign suggests net avoidance

Life cycle coal consumption (MJ)	Physical	Economic	OLS
$r_{\text{electricity}}$	4.08E-01	3.22E+00	−8.80E-03
r_{heat}	4.11E-02	1.21E+00	1.60E-02
$r_{\text{white clay}}$	−1.10E-03	9.52E-03	−3.85E-03
$r_{\text{red clay}}$	−1.10E-03	8.65E-03	−1.38E-03
$r_{\text{recycled inerts}}$	−2.61E-03	2.24E-02	3.28E-04
r_{sand}	−3.07E-03	2.64E-02	−7.24E-03
r_{gravel}	−3.07E-03	1.93E-02	−6.51E-03
$r_{\text{olive cake}}$	−6.08E-05	1.02E-02	−9.80E-04
r_{straw}	−1.52E-05	4.12E-03	−3.44E-04
$r_{\text{crude oil}}$	2.02E-03	5.91E-02	7.79E-04
$r_{\text{diesel oil}}$	3.26E-03	3.87E-02	1.88E-04
$r_{\text{fuel oil}}$	3.26E-03	1.52E+00	−3.83E-02
$r_{\text{natural gas}}$	1.20E-01	1.46E+00	7.58E-03
r_{bricks}	8.56E+02	1.22E+04	4.75E-06

respective intensity matrices.) Since coal is listed as a resource used by most of the given processes when operated at positive scale, the final intensities for coal are expected to be negative indicating resource consumption. Only a few of the products match this expectation under physical and OLS allocation, and none of the products match expectation under economic allocation. The positive intensities seem to suggest that these products somehow create or contribute to the avoided use of coal elsewhere. Furthermore, under the economic allocation, all of the products system wide appear to produce (or avoid the use of) net amounts of coal despite the fact that it is a direct input into seven of the eight processes. Finally, the combined life cycle consumption of coal by straw and olive cake ($r_{\text{straw}}+r_{\text{olive cake}}$) under OLS allocation is smaller than the direct consumption of coal into the whole biomass process (Marvuglia et al. 2010, Table 2). Furthermore, the biomass depends on the use of diesel oil as an input in the agricultural phase. Both processes must be operated to provide necessary inputs of biomass for the brick process, and yet, under OLS assumptions, life cycle consumption of coal by both types of biomass is smaller than the coal consumption of the agricultural phase.

3.2.2 Least squares equivalent partitioned matrices

Reducing the number of products of the original data to the same dimension by instituting the recovery and substitution of inerts for sand, one can determine the set of allocation weights consistent with OLS allocation for a given vector of emissions, b, by reconciling Eqs. 9 and 10. That is, one solves the equation

$$\left(A_*^T\right)^{-1} b_* = \left(A^T\right)^+ b \tag{17}$$

for the allocation weights which appear in $\boldsymbol{A}_*$ and $\boldsymbol{b}_*$. In Table 3, the allocation weights for selected interventions are compared to the weights for the traditional allocations, as derived from information in (Marvuglia et al. 2010, Tables 3 and 4). As crude oil, natural gas, and brick (after assuming substitution of inerts for sand) were derived from single function processes, they are omitted from the table. For OLS allocation, two other interventions are presented along with coal to illustrate that implied weights may differ for each of the distinct interventions, as if a different allocation principle has been chosen in each case. Note that in all outcomes, the weights assigned to a pair of products sum to 1; however, some of the weights assigned to joint products by ordinary least squares are clearly indefensible in the sense that some inputs may become outputs and while other products may be assigned a share of inputs greater than 100 % in the corresponding technology and interventions matrices.

Both DLS and TLS derive their inventories from a perturbed version of Eq. 1. For each type of intervention, DLS and TLS will yield a scalar intensity associated with brick r_{brick}. Fixing this value still leaves five degrees of freedom in a solution of Eq. 8. One could find infinitely many intensity vectors in which the component r_{brick} is equal to DLS or TLS outcome, perhaps even with all positive weights. For example, one possible family of TLS-equivalent partitioning allocations for coal could be obtained by setting ($w_{\text{white clay}}$, w_{sand}, $w_{\text{olive cake}}$)=(0.5, 0.5, 0.8) and searching the level curve r_{brick}=250 MJ *coal* of the function shown in Fig. 3 for corresponding ($w_{\text{electricity}}$, $w_{\text{diesel oil}}$) pairs.

The set of all solutions to the underdetermined system of equations Eq. 8 is infinite, and each allocation, least squares techniques now included, merely selects one particular solution and interprets the coordinate of interest in the solution vector. The utility *of any of these allocation methods* is questionable if no further thought is given to the side conditions associated with them. When acquisition of additional data is simply impossible, any statement as to the quality or correctness of a particular inventory after allocation must be tied to just the belief that the side conditions implied by the partitioning criterion are at least plausible. It is imperative that a decision maker should at least try to understand what those implications are!

3.3 Discrepancy vectors

The reported discrepancy vectors for economic and physical allocations (both invertible systems) were given along with

Table 3 Description of stated or implied weights assigned to each pair of coproducts. Weights for economic and physical allocation were as stated in (Marvuglia et al. 2010, Tables 3 and 4). In OLS allocation, all pairs of coproduct weights sum to 1, but some weights are negative or exceed 1

Unallocated process	Individual products	Stated or implied allocation weights				
		Economic	Physical	OLS MJ geotherm. energy	OLS kg COD	OLS, MJ coal
Electricity and heat	MJ of electricity	0.518	0.8	−0.375	0.206	−0.289
	MJ of heat	0.482	0.2	1.375	0.794	1.289
Clay	kg of white clay	0.524	0.5	1.600	−0.921	0.516
	kg of red clay	0.476	0.5	−16.180	312.459	0.262
Sand and gravel	kg of gravel	0.577	0.5	17.180	−311.459	0.738
	kg of olive cake	0.423	0.5	−0.600	1.921	0.484
Supply of biomass	kg of sand	0.713	0.8	−5.515	0.494	0.742
	kg of straw	0.287	0.2	6.515	0.506	0.258
Oil derivatives	MJ of diesel oil	0.058	0.5	−0.004	0.017	−0.005
	MJ of fuel oil	0.942	0.5	1.004	0.983	1.005

discrepancy vectors from each of the three regression-based methods in (Marvuglia et al. 2010, Table 12). Observing that residuals among the regression-based techniques were of much greater magnitude than either of the traditional allocation methods, the following conclusion was reached in (Marvuglia et al. 2010, p. 1028):

> "On the basis of the observation of the discrepancy vectors alone, the traditional allocation based solutions may therefore appear to be preferred to the solutions obtained using the regression techniques. However, a complete judgment should also take into account the observation of the differences among the inventory vectors since the inventory vector is the nal goal of the LCI phase."

The discrepancies which were tabulated provide vital information and merit more thorough examination, however.

Comparison of the residuals of the traditional partitioning allocations to those of any of the least squares techniques does not make any sense, since the former results due to machine precision in the numerical solution of an invertible system of equations while the latter is due to the lack of a unique solution and the choice of an approximate solution of the overdetermined system Eq. 1. The tabulated residuals ((Marvuglia et al. 2010, Table 12)) associated with OLS, DLS, and TLS correspond to the quantities $As_{\text{OLS}}-f_{\text{brick}}$, $\Delta A_{DLS}s_{DLS}$, and $\Delta A_{\text{TLS}}s_{\text{TLS}}{}^{-}\Delta f_{TLS}$, respectively. Beyond any residuals associated with the functional unit in OLS and TLS, the residuals of the technology matrix in DLS and TLS must be carefully

Table 4 Percentage change ($100 * \Delta A_{\text{DLS}} /|A|$) in cells of A implied by $\Delta \text{A}_{\text{DLS}}$

	Electricity and heat	Clay (red and white)	Sand and gravel	Crude oil	Oil derivatives	Natural gas	Supply of biomass	Bricks and inerts	Functional unit
MJ of electricity		95.6	14.1	−	−99.2	−	+		−
MJ of heat		−	−	**253.5**	3.7	3.0	−		+
kg of white clay	−		+	−	−	−	+		−
kg of red clay	+		−	+	+	+	−		+
kg of recycled inerts	+	−	−	+	+	+	−		+
kg of sand	−	+		−	−	−	+		−
kg of gravel	+	−		+	+	+	−		+
kg of olive cake	−	+	+	−	−	−			−
kg of straw	+	−	−	+	+	+			+
MJ of crude oil	−	+	+	−18.3	−2.7	−	+	+	−
MJ of diesel oil	+	−2.2		26.3					+
MJ of fuel oil	−	+	+	−	−12.1	−	+		−
MJ of natural gas		−	−	+	+		−		+
Ton of bricks									

Changes to structural zeros are indicated by +/-. Numeric values indicate a percentage change greater than +/- 2% of the original cell value. Bold face denotes a flow which has changed direction after revision

examined. Tables 4 and 5 examine these adjustments to the data. While small in the sense that the perturbations minimize the Frobenius norm of the correction matrix $[\boldsymbol{\Delta A}|\boldsymbol{\Delta f}]$, the adjusted entries potentially:

- change known zeros to small amounts of inputs or outputs,
- indicate large changes for some element, perhaps even changing sign so that inputs become outputs or vice versa, and
- correspond to a very different technology matrix than the one intended.

In Tables 4 and 5, the corrections matrix $[\boldsymbol{\Delta A}|\boldsymbol{\Delta f}]$ for DLS and TLS are expressed as percentage changes are given for non-zero elements of the original technology matrix. The two tables differ primarily in the bottom row and the functional unit column since TLS permits changes to the stated functional unit as well as the technology matrix. In each table, a blank cell indicates a change smaller than 2 % (in absolute value) of the cell's original value. The signs +/− indicate changes of structural zeros into small amounts of outputs or inputs, respectively. After perturbation, the brick production unit appears to generate crude oil, changes in six cells were greater than 10 % of the magnitude of the original cell entry, and one cell entailed a 250 % change and a reversal of sign, indicating that, in the perturbed technology matrix $\boldsymbol{A}+\boldsymbol{\Delta A}$, heat became an output of the crude oil process rather than an input used to do work and pump oil. This is another manifestation of the shortfall in system-wide energy inputs mentioned in Section 3.1. Consequently, the scale factors derived by DLS and TLS methods correspond to technology matrices with little resemblance to the original data.

Note that the perturbations in $\boldsymbol{\Delta A}$ also depend on the stated functional unit. Consequently, a researcher could perform DLS or TLS on a given data matrix for two different functional units and derive rather different sets of corrections to the technology matrix. As stated in Marvuglia et al. (2010), it is possible that the recorded data were measured with some error, but surely the extent of the measurement errors should be independent of which functional unit is stated for study.

In summary, Marvuglia et al. appeared to favor the TLS method on the grounds that it (1) produced (a few) results that were similar to physical allocation and (2) on the basis of asymptotic properties of TLS parameter estimates (i.e., strong consistency of TLS estimators). Regarding the first point, as discussed in the previous section, one should not necessarily expect agreement among inventories derived under different constraints. Among the five tabulated inventories, one can find agreement in some components of the inventories and radical disagreement in others, leaving the issue of which value(s) seem most reasonable up for debate. The argument for the second point simply appears to be that TLS will yield better results on bigger data sets. This reasoning is neither justified theoretically, nor does it resolve the allocation problem in any way. To act on the TLS scale factors, one must believe that the changes to the technology matrix $\boldsymbol{A}+\boldsymbol{\Delta A}_{\mathbf{TLS}}$ are plausible. These perturbations may be quite significant as they could entail changes to structural zeros and perhaps changes of inputs into outputs or vice versa. In any case, the asymptotic properties

Table 5 Percentage change ($100 * \Delta A_{\mathrm{TLS}} /|A|$) in cells of A implied by ΔA_{TLS}. The same sign conventions used in Table 4 apply. DLS and TLS differ primarily by perturbations in the final row of the technology matrix. By assumption, TLS permits small discrepancies in the stated functional unit; small changes are also introduced in the final row to compensate for error in the stated functional unit

	Electricity and heat	Clay (red and white)	Sand and gravel	Crude oil	Oil derivatives	Natural gas	Supply of biomass	Bricks and inerts	Functional unit
MJ of electricity		95.6	14.1	−	−99.2	−	+		−
MJ of heat		−	−	**253.5**	3.7	3.0	−		+
kg of white clay	−		+	−	−	−	+		−
kg of red clay	+		−	+	+	+	−		+
kg of recycled inerts	+	−	−	+	+	+	−		+
kg of sand	−	+		−	−	−	+		−
kg of gravel	+	−		+	+	+	−		+
kg of olive cake	−	+	+	−	−	−	−		−
kg of straw	+	−	−	+	+	+			+
MJ of crude oil	−	+	+	−18.3	−2.7	−	+	+	−
MJ of diesel oil	+	−2.2		26.3					+
MJ of fuel oil	−	+	+	−	−12.1	−	+		−
MJ of natural gas		−	−	+	+		−		+
Ton of bricks	−	+	+	−	−	−	+		−

Changes to structural zeros are indicated by +/-. Numeric values indicate a percentage change greater than +/- 2% of the original cell value. Bold face denotes a flow which has changed direction after revision

of s_{TLS} do not guarantee that the discrepancies are small in any practical sense. The issue of allocation is one of rank deficiency or non-identifiability of the inventory of interest. In fact, applying TLS on a data set of the size of the entire unallocated ecoinvent v1.3 database with 2,632 rows (products) and 2,471 columns (processes) (Marvuglia et al. 2010, p. 1021) would not overcome the rank deficiency, *requiring some 161 additional constraints*, in any meaningful way.

4 Conclusions

In actuality, the fundamental problem of allocation is one of too few linear equations and too many unknowns. The components of the intensity matrix to be interpreted in the inventory step may not be uniquely identified in the given data. In such cases, the final inventory value will be *determined* by the additional information brought to bear on a solution of Eq. 8.

Upon reexamination of the brick production data, issues of data quality and the representativeness of included processes may explain some of the unusual disparities in the originally provided inventories. Each allocation by least squares techniques or by more traditional means of partitioning will yield a result irrespective of data quality. Consequently, it is important to define what constitutes a legitimate technology matrix and to identify and correct any gross errors before carrying out the remaining steps of the life cycle analysis.

Of more general concern, an inventory produced by partitioning must necessarily bring some additional information to bear on the problem. In that sense, the least squares techniques are no different than more traditional partitioning allocations. Inventories generated by least squares techniques are, in fact, equivalent to some partitioning of the technology and interventions matrices. We caution practitioners that the use of least squares techniques for the purposes of allocation is not a black-box solution to the problem of allocation and may not generate an output that is any more physically meaningful or actionable than already existing methods of partitioning.

A square, invertible technology matrix obtained through allocation often marks the starting point for structural path analysis and many existing methods of uncertainty propagation in LCA. Like the brick example, sensitivity of the inventory to allocation assumptions is often examined through application of several allocations to the same data. The *mathematical structure* of the allocation problem permits more general and thorough understanding. In particular, forthcoming work will provide ways to address the sensitivity of specific comparisons of products to allocation assumptions Hanes et al. (Forthcoming) and describe conditions for the existence of inverse-positive outcomes Cruze (2013) which may facilitate the exploration of the set of all allocations.

To satisfactorily avoid partitioning, practitioners must either (1) acquire more data on similar processes involved in the life cycle of the target product or upstream inputs or (2) state additional assumptions about life cycle intensities of select products within the system more judiciously. As it stands, practitioners can follow prescribed recipes for allocation without regard to the veracity of additional constraints, much to the detriment of the inventory and subsequent impact assessment and interpretation phases of any life cycle assessment.

References

Cherubini F, Strømman AH (2011) Life cycle assessment of bioenergy systems: state of the art and future challenges. Bioresour Technol 102(437):451

Cruze N, Goel PK, Bakshi BR (2013a) On the rigorous proof of fuzzy error propagation with matrix-based LCI. Int J Life Cycle Assess 18(516):519

Cruze, N. B. (2013). Addressing Allocation and Disparity in Methods of Life Cycle Inventory. Ph. D. thesis, Ohio State University

Cruze, N. B., P. K. Goel, and B. R. Bakshi (2013b). Allocation in life cycle inventory: Partial set of solutions to an ill-posed problem. Int. J. Life Cycle Assess. Accepted

DeGroat R, Dowling E (1993) The data least squares problem and channel equaliza-tion. IEEE Trans Signal Process 41(1):407–411

Golub G, Kahan W (1965) Calculating the singular values and pseudo-inverse of a matrix. J Soc Ind Appl Math B Met 2(205):224

Hanes, R. J., N. B. Cruze, P. K. Goel, and B. R. Bakshi. The allocation games. Environ Sci Technol. Submitted

Heijungs R, Suh S (2002) The computational structure of life cycle assessment. Kluwer Academic Press, Dordrecht

Heijungs R, Tan R (2010) Rigorous proof of fuzzy error propagation with matrix-based LCI. Int J Life Cycle Assess 15(1014):1019

Marvuglia A, Cellura M, Heijungs R (2010) Toward a solution of allocation in life cycle inventories: the use of least-squares techniques. Int J Life Cycle Assess 15(1020):1040

Marvuglia A, Cellura M, Pucci M (2012) A generalization of the orthogonal regression technique for life cycle inventory. Int J Agr Environ Inf Syst 3(51):71

Peters G (2007) Efficient algorithms for life cycle assessment, input-output analysis, and Monte-Carlo analysis. Int J Life Cycle Assess 12(373):380

Strømman A, Peters G, Hertwich E (2009) Approaches to correct for double counting in tiered hybrid life cycle inventories. J Clean Prod 17(248):254

Tan R (2008) Using fuzzy numbers to propagate uncertainty in matrix-based LCI. Int J Life Cycle Assess 13(585):592

Van Huffel, S. (Ed.) (1997). Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling. Society for Industrial and Applied Mathematics

Van Huffel, S. and J. Vandewalle (Eds.) (1991). The Total Least Squares Problem: Compu-tational Aspects and Analysis. Society for Industrial and Applied Mathematics

Wardenaar T, van Ruijven T, Mendoza Beltran A, Vad K, Guinée J, Heijungs R (2012) Differences between LCA for analysis and LCA for policy: a case study on the consequences of allocation choices in bio-energy policies. Int J Life Cycle Assess 17(1059):1067